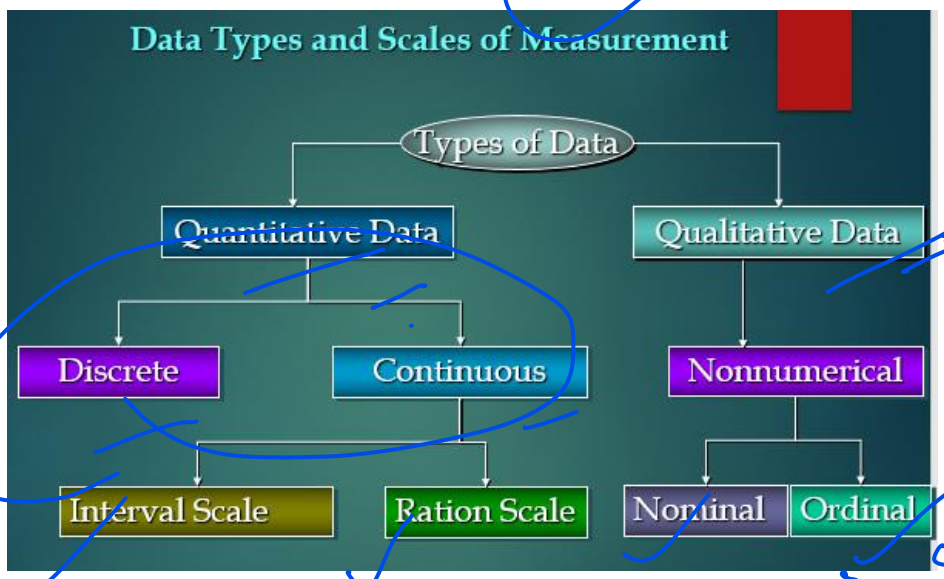
UNIT-II STATISTICS IN PSYCHOLOGY

Types of Scale

Measures of Central Tendency and Dispersion.



Types of scale

1. Nominal Scale
2. Ordinal Scale
3. Interval Scale
4. Ratio Scale

1. Nominal Scale

In nominal scales are used to name or identify people, objects, or events.

It is used to represent qualitative data.

If the data has only two distinct values, then it is called as Binary/ Dichotomous

SHOUT
SHARE
IT

Example- Gender (male and female), Blood (A, B, AB,)

2. Ordinal Scale

In ordinal scales, numbers represent rank order and indicate the order of quality or quantity, but they do not provide an amount of quantity or degree of quality.

Here a natural order exists among the groups, but the difference between the groups is not necessarily equal.

It is used to represent qualitative data.

Example- Educational Qualification

3. Interval Scale.

Ordinal Scale is defined as a variable measurement scale used to simply depict the order of variables and not the difference between each of the variables.

Origin of this scale is absent due to which there is no fixed start or "true zero".

Number '0' is meaningless

If zero is used, it simply serves as a reference point on the scale but does not indicate the complete absence of the characteristic being measured

It is used to compare (on the basis of differences) Quantitative data.

Example include temperature measured in centigrade, time (even though oooohrs exist, there is nothing like no time)

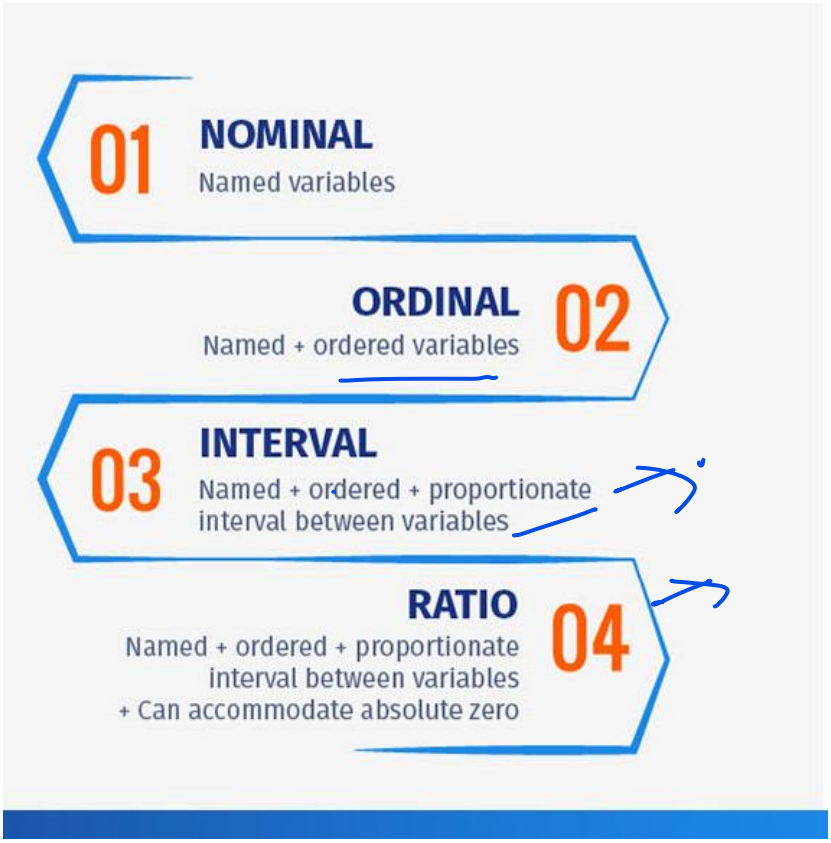
4. **Ratio Scale**

Ratio scales have all of the characteristics of interval scales as well as a true zero, which refers to complete absence of the characteristic being measured. Physical characteristics of persons and objects can be measured with ratio scales.

It is used to measure quantitative data (on the basis of fraction)

Number '0' is meaningful

Example – Height, revenue, cost, profit



By Dr. Farah
Telegram Id- https://t.me/ugcnetpsychologyforall

Difference Between Interval and Ratio Scale

Interval Scale	Ratio Scale
Does not have true zero	Has true zero.
Example- Temperature,	Example- money, height weight, BP
Negative measurement (-1, -2 temp)	Positive measurement
Does not measure complete absence	Measures complete absence
Zero is meaningless	Zero is meaningful

Central Tendency

"A sort of average or typical value of the items in the series and its function is to summarize the series in terms of its average value." (Tate, 1955)

Measures of central tendency-

- Mean
- Median
- Mode

Mean (M)

Mean is the simplest and most useful measure of central tendency.

The mean is calculated by adding up all of the scores and dividing by the number of individual scores. We can write this in equation form as follows:

$$\overline{X} = \frac{\sum X}{N}$$

(X is the statistical symbol for a score, ∑ is the summation or sigma sign, ∑X means add up all of the scores, X N is the number of scores, X is the statistical symbol for the arithmetic mean of a set of scores

Example- 5, 10, 8, 6, 5, 4 (ungrouped data)

Calculation of mean for grouped data

To calculate the mean of grouped data, the first step is to determine the midpoint (also called a class mark) of each interval, or class. These midpoints must then be multiplied by the frequencies of the corresponding classes. The sum of the products divided by the total number of values will be the value of the mean.

Example –

Marks	Students (f)	X (midpoint)	fX
51 – 55	2	53	106
56 – 60	7	58	406
61 – 65	8	63	504
66 – 70	4	68	272
Totals	21		1288

where X is the midpoint of the class and f is the frequency.
As a result, the formula

mean= 1288/21
61.33

$$\bar{X} = \frac{\sum fX}{n}$$

while calculating and interpreting the mean we need to make two assumptions of our data-

The distribution is symmetrical. This means that there is not much skew, and no outliers on one side. (You can still calculate the mean if the distribution is not symmetrical, but interpreting it will be difficult because the result will give you a misleading value.)

2. The data are measured at the interval or ratio level. We saw earlier that data can be measured at a number of levels. It would not be sensible to calculate the mean of bags' colour. If we observed that half of the people have blue bags, and half of the people have yellow bags, it would make no sense to say that the mean (average) bag colour was green.

Median

The second most common measure of central tendency is the median. It is the middle score in a set of scores. The median is used when the mean is not valid, which might be because the data are not symmetrically or normally distributed, or because the data are measured at an ordinal level.

This midpoint value is the point at which half the observations are above the value and half the observations are below the value. The median is determined by ranking the observations and finding the observation that are at the number [N + 1] / 2 in the ranked order.

Median for Ungrouped data-

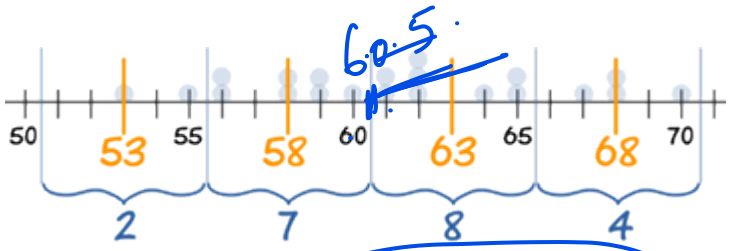
7, 9, 10, 12, **13**, 14, 17, 18, 19 (uneven)

7, 9, 10, 12, 13, 14, 17, 18, 19, 20 (even)

By Dr. Farah
Telegram Id- https://t.me/ugcnetpsychologyforall

Median For Grouped data

Marks	Students (f)	Cf
51 - 55	2	2
56 - 60	7	9
61 - 65	8	17
66 - 70	4	21
Totals	21	39



$$\text{Median} = l + \frac{\left(\frac{N}{2} - m\right)}{f} \times c$$

l= lower class of the group having median

N=total number of values

m=cumulative frequency before the median group

c= width of class interval

f=frequency of the median

$$60.5 + \frac{(21/2) - 9}{8} \times 5$$

$$= 60.5 + 0.9375$$

$$= \mathbf{61.4375}$$

Mode

The final measure of central tendency, rarely reported in research, is the mode. It is the most frequent score in the distribution or the most common observation among a group of scores. The mode is the best measure of central tendency for categorical data (although it's not even very useful for that).

The problem with the mode is that it doesn't tell you very much. For example, imagine that we have a group of 45 males and 20 females. What is the mode? The mode is male. But that just tells us that there were more males than females. It doesn't take much more space to say that there were 45 males and 20 females.

Mode for Ungrouped data

25, 29, 24, 25, 27, 25, 28, 25, 29

Mode for grouped Data

Marks	Students (f)
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4
Totals	21

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h$$

where l = lower limit of the modal class,

h = size of the class interval (assuming all class sizes to be equal),

f_1 = frequency of the modal class,

f_0 = frequency of the class preceding the modal class,

f_2 = frequency of the class succeeding the modal class.

$$60.5 + \frac{(8-7)}{(8-7)(8-4)} \times 5$$

$$= 60.5 + (1/5) \times 5$$

$$= \mathbf{61.5}$$

Mode= 3 median- 2mean

Mo = 3 x 61.43- 2 x 61.33

Mo = 184.29 – 122.66

61.63

Comparison of Mean, Median Mode

When deciding which of the three measures of central tendency to use, you should take into account the distribution of the scores. When the distribution is unimodal (i.e., has one mode) and symmetrical, then the mode, median and mean will have very similar values.

The mean is often the best average, for a couple of reasons. First, unlike the median, it uses all of the information available. Every number in the dataset has some influence on the mean. Secondly, the mean also has useful distributional properties (don't worry about what this means, we will cover it later), which the median does not have. The downfall of the mean is that it is affected by skew and by outliers.

Psychology	Geography
17	20
19	20
19	20
17	20
18	1
Mean=18	Mean=16.2

Scale Type	Permissible Statistics	
Nominal	Mode	Chi square test
Ordinal	Mode/Median	
Interval	Mean, Standard Deviation	Correlation, Regression, t Test, Anova
Ratio		

Measures of Dispersion and Spread

1. Range
2. Inter-Quartile Range
3. Standard Deviation

Range

The range is the simplest measure of dispersion. It is simply the distance between the highest score and the lowest score. It can be expressed as a single number, or sometimes it is expressed as the highest and lowest scores.

Example- Variable range – 4, 11, 17, 12, 3, 15, 10, 2, 8,

To find the range we find the lowest value (2) and the highest value (17).

Sometimes the range is expressed as a single figure, calculated as:

Range = highest value – lowest value, sometimes it is expressed as the range of scores: here the range is from 2 to 17.

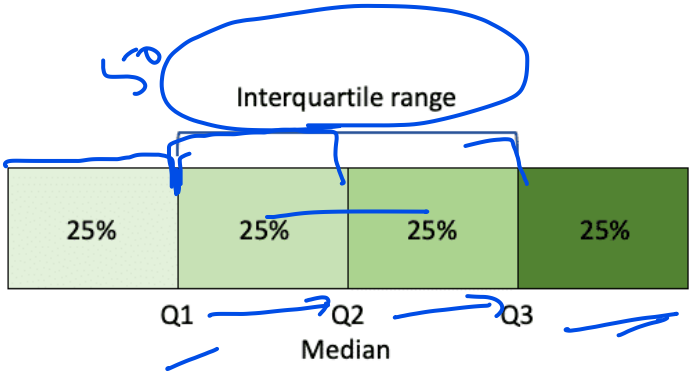
Example- For example 'the ages of the participants ranged from 18 years 21 years 23 years 18 years 19 years 19 years 19 years 33 years 18 years 19 years 19 years 20 years

Note-One trouble with range is that it can be heavily influenced by extreme cases. Thus the 33-year-old participants in the above set of ages is having a big influence on the range of ages of the participants. This is because he or she is much older than most of the students. For this reason, the interquartile range has advantages.

Inter-quartile Range

The inter-quartile range (IQR) is used with ordinal data or with non-normal distributions. If you use the median as the measure of central tendency then you'll probably use the inter-quartile range as a measure of dispersion.

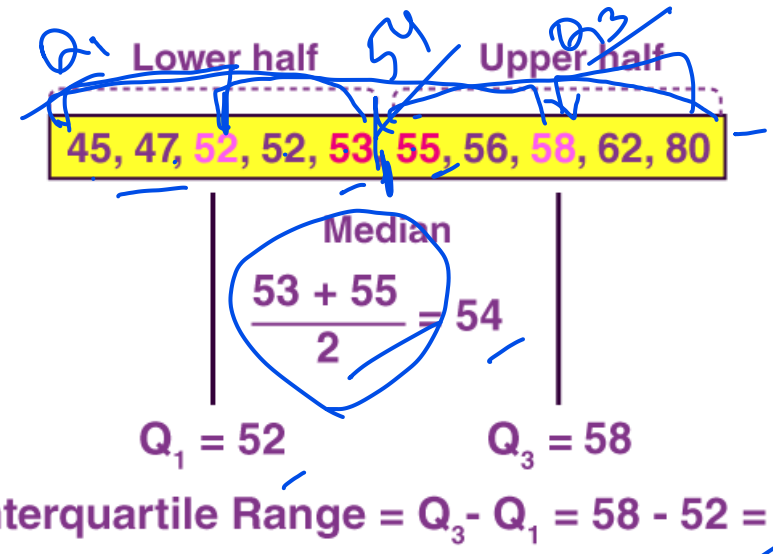
To calculate the interquartile range, we split the age distribution into quarters and take the range of the middle two quarters (or middle 50%), ignoring the extreme quarters. Since we have 12 students, we delete the three youngest (the three 18-year-olds) and the three oldest (aged 33, 23 and 21). This leaves us with the middle two quarters (the middle 50%) which includes five 19-year-olds and one 20-year-old. The range of this middle two quarters, or the interquartile range, is one year (from 19 years to 20 years).



In other words,

count	Value	
1	18	Low Quartile- 19
2	18	
3	18	
4	19	
5	19	Mid-point Between 6-7
6	19	
7	19	
8	19	

9	20	High Quartile- 20
10	21	
11	23	
12	33	



The interquartile range is sometimes a better indicator of the variability of, say, age than the full range because extreme ages are excluded.

Useful as the range is, a lot of information is ignored. It does not take into account all of the scores in the set, merely the extreme ones. For this reason, measures of spread or variability have been developed which include the extent to which each of the scores in the set differs from the mean score of the set.